

## Übung Multivariate Analyseverfahren – SoSe 2017

Dr. Gerrit Bauer

### Übungsblatt 1 – Abgabe bis 17.05.2017

Wir stellen einen Paneldatensatz für die beiden Jahre 1992 und 2010 aus den Originaldaten des SOEP zusammen, mit dem wir die Zeitverwendung für Hausarbeit analysieren können.

1. Erstellen Sie ein Do-file „03\_Übungsblatt1“. Legen Sie Ihre Arbeitsverzeichnisse fest, in dem Sie globale Macros `pfad1` und `pfad2` definieren (Befehl `global`). (`pfad1` ist der Ordner, in dem die SOEP-Originaldaten liegen; `pfad2` der Ordner, in dem Sie Ihre rekodierten Daten speichern wollen.)
2. Finden Sie die Variablen zum Zeitaufwand für Hausarbeit (in Std.) an einem typischen Werktag für die Wellen 1992 und 2010 in den Personendaten `$p.dta` des SOEP. Matchen Sie die Information zur Hausarbeit aus beiden Jahren und speichern Sie die Info als neuen Datensatz `hw1.dta` in `pfad2`.

(Hinweis: Erzeugen Sie zunächst für jedes der beiden Jahre einen Querschnittsdatsatz, der nur die Personennummer, das Befragungsjahr und die Variable zur Hausarbeit enthält. Nennen Sie die Variablen: `id`, `year`, `hw`. Matchen Sie dann die Infos aus beiden Jahren in einem Paneldatensatz.)

3. Erstellen Sie einen Datensatz `hw2.dta`, der für die beiden Jahre nur die folgenden Variablen enthält:
  - Personennummer (neuer Variablenname: `id`)
  - Befragungsjahr (`year`)
  - Aktuelle Haushaltsnummer (`hhnrakt`)
  - Personennummer PartnerIn (`pid`)
  - Familienstand (`famstd`)
  - Bruttomonatsverdienst in Euro (`labgro`)

Verwenden Sie generierte Variablen. Speichern Sie den neuen Datensatz (in `pfad2`).

4. Erstellen Sie einen Datensatz `hw3.dta` (in `pfad2`), der für jede Person nur die folgenden zeitkonstanten Variablen enthält:
  - Personennummer (`id`)
  - Geschlecht (`sex`)
  - Geburtsjahr (`gebjahr`)

Verwenden Sie die Infos aus dem Datensatz `ppfad.dta`.

5. Spielen Sie jetzt die Daten aus `hw1.dta`, ..., `hw3.dta` der Reihe nach zusammen (Befehl `merge`). Betrachten Sie dazu in jedem Schritt zuerst die Datenstruktur der beiden Datensätze, die sie matchen wollen (`master` und `using`). Verwenden Sie den Befehl `list` (oder `browse`), um sich immer die ersten 20 Beobachtungen ausgeben zu lassen. Überlegen Sie sich dann, wie der `merge` Befehl lauten muss.
6. Speichern Sie den resultierenden Datensatz als `housework_92_10.dta` (in `pfad2`).
7. (\*) Erstellen Sie einen Paneldatensatz `hw4.dta`, der für die beiden Jahre und für jeden Haushalt nur die folgenden Variablen enthält:
  - Aktuelle Haushaltsnummer (`hhnrakt`)
  - Befragungsjahr (`year`)
  - Wohnort Ost-/Westdeutschland (`east`)
  - (Korrigiertes) Netto-Haushaltseinkommen (`hhinc`)Verwenden Sie generierte Variablen (Originaldaten `$hgen.dta`). Matchen Sie die Daten `hw4.dta` mit dem Datensatz `housework_92_10.dta`.
8. (\*) Lassen Sie sich deskriptive Statistiken (Mittelwerte bzw. Anteile) für die Variablen im Datensatz ausgeben, getrennt nach Geschlecht und Erhebungsjahr. (Dazu müssen Sie zuerst Missings definieren.) Wie hat sich der Zeitaufwand von Frauen und Männern entwickelt?
9. Schicken Sie Ihr Do-file an [gerrit.bauer@lmu.de](mailto:gerrit.bauer@lmu.de). Geben Sie im Betreff der email das Kürzel „MVSO17IhrNachname“ an.

**Aufgaben mit (\*) sind Zusatzaufgaben.**

## Übung Multivariate Analyseverfahren – SoSe 2017

Dr. Gerrit Bauer

### Übungsblatt 2 – Abgabe bis 07.06.2017

Wir stellen einen Paneldatensatz für die Jahre 1992 bis 2010 aus den Originaldaten des SOEP zusammen, mit dem wir die Zeitverwendung für Hausarbeit analysieren können. Erweitern Sie das Do-file zu Übungsblatt 1 (03\_Übungsblatt1.do). Erstellen Sie ein neues Do-file (04\_Übungsblatt2.do). Verwenden Sie Schleifen, um alle Jahre 1992,..., 2010 zu matchen.

1. Legen Sie Ihre Arbeitsverzeichnisse fest, in dem Sie globale Macros `pfad1` und `pfad2` definieren. (`pfad1` ist der Ordner, in dem die SOEP-Originaldaten liegen; `pfad2` der Ordner, in dem Sie Ihre rekodierten Daten speichern wollen.)
2. Matchen Sie die Information zur Hausarbeit (in Std., typischer Werktag) aus allen Jahren 1992,..., 2010 und speichern Sie Personennummer (`id`), Befragungsjahr (`year`) und die Variable zur Hausarbeit (`hw`) im neuen Datensatz `hw1.dta`. (Hinweis: Variablen zum Zeitaufwand für Hausarbeit in den Personendaten `$p.dta`:

```
ip0103 jp0203 kp0803 lp0203 mp0203 np0203 op0503 ///
pp02a3 qp05a3 rp0207 sp1103 tp1007 up0203 vp0207 ///
wp6203 xp0207 yp1203 zp0203 bap0303
```

3. Erstellen Sie einen Datensatz `hw2.dta` aus (`$pgen`), der für die Jahre 1992,..., 2010 nur die folgenden Variablen enthält:
  - Personennummer (neuer Variablenname: `id`)
  - Befragungsjahr (`year`)
  - Aktuelle Haushaltsnummer (`hhnrakt`)
  - Personennummer PartnerIn (`pid`)
  - Familienstand (`famstd`)
  - Bruttomonatsverdienst in Euro (`labgro`)
  - Erwerbsstatus (`lfs`) (**neu**)
4. Erstellen Sie einen Datensatz `hw3.dta` (aus `ppfad.dta`), der für jede Person die folgenden zeitkonstanten Variablen enthält:
  - Personennummer (`id`)
  - Geschlecht (`sex`)
  - Geburtsjahr (`gebjahr`)
  - Teilstichprobe (`psample`) (**neu**)

Der Datensatz soll zusätzlich die zeitveränderliche Variable zum Wohnort in Ost-/Westdeutschland enthalten (`east`). Erstellen Sie diese Variable aus den Variablen `$sampreg`.

(Hinweis: Benennen Sie die Variablen `isampreg`, ..., `basampreg` in einer Schleife um in `east1992`, ..., `east2010`. Verwenden Sie dann den Befehl `reshape`, um den Datensatz ins Long Format zu bringen.)

5. Erstellen Sie einen Paneldatensatz `hw4.dta` aus (`$hgen`), der für die Jahre 1992,..., 2010 und für jeden Haushalt nur die folgenden Variablen enthält:
  - Aktuelle Haushaltsnummer (`hhnrakt`)
  - Befragungsjahr (`year`)
  - (Korrigiertes) Netto-Haushaltseinkommen (`hhinc`)
6. Erstellen Sie einen Paneldatensatz `hw5.dta` aus (`$pequiv`), der für die Jahre 1992,..., 2010 nur die folgenden Variablen enthält:
  - Personennummer (`id`)
  - Anzahl HH-Mitglieder im Alter von 0-14 J. (`kids14`)
  - Befragungsjahr (`year`)
  - Region Wohnort Ost-/Westdeutschland (`east2`)
7. Spielen Sie jetzt die Daten aus `hw1.dta`, ..., `hw5.dta` der Reihe nach zusammen. Speichern Sie die Daten als `housework.dta`.
8. (\*) Erstellen Sie einen (dyadischen) Paneldatensatz `housework_dyad.dta` für heterosexuelle Paare. Gehen Sie aus von `housework.dta`.
  - Erstellen Sie einen Datensatz `women.dta`, der nur Frauen enthält, die mit einem/r Partner/in zusammen leben (`pid>0`).
  - Erstellen Sie einen Datensatz `men.dta`, der nur Männer enthält, die mit einem/r Partner/in zusammen leben.  
Der Datensatz soll die Variablen `pid` `hw` `gebjahr` `sex` `labgro` `famstd` `lfs` enthalten.
  - Matchen Sie die Personenjahre aus `women.dta` und `men.dta` mit `merge`.  
(Hinweis: Damit sie die Daten matchen können, müssen Sie in `men.dta` zuerst die Personennummer `pid` des Partners in `id` umbenennen. Außerdem müssen Sie die Variablen `hw`, ..., `lfs` umbenennen; z.B. könnten Sie in einer Schleife die Endung `_m` an die Variablennamen anhängen.)
9. Schicken Sie Ihr Do-file an [gerrit.bauer@lmu.de](mailto:gerrit.bauer@lmu.de). Geben Sie im Betreff der email das Kürzel „MVSO17IhrNachname“ an.

**Aufgaben mit (\*) sind Zusatzaufgaben.**

## Übung Multivariate Analyseverfahren – SoSe 2017

Dr. Gerrit Bauer

### Übungsblatt 3 – Abgabe bis 21.06.2017

Wir arbeiten mit dem Paneldatensatz `housework.dta`, den wir mit Übungsblatt 2 erstellt haben.

1. Prüfen Sie die Konsistenz der Variablen `marstat` im Längsschnitt. Identifizieren Sie Personen, die in (mindestens) einer Welle verheiratet, getrennt/geschieden oder verwitwet sind und in (mindestens) einer späteren Welle ledig. Gibt es Personen mit inkonsistenten Angaben in den Daten? Wie viele?
2. Generieren Sie (a) eine Variable `pys`, die für jede Person im Datensatz die Anzahl der Personenjahre zählt; (b) eine Variable `pys2`, die für jede Person im Datensatz die Anzahl der Personenjahre mit gültigen Werten auf den Variablen `hw labgro marstat hhinc age` zählt. Für wie viele Personen beobachten wir die Variablen in mindestens 2 Personenjahren? Wie viele Personenjahre verbleiben mit der Restriktion in (b)?
3. Wenden Sie folgende within-Transformationen auf die Daten an (alle Variablen aus 2.):
  - (a) Demeaning
  - (b) First-Differencing

Rechnen Sie bivariate (Pooled) OLS Regressionen mit den nicht transformierten Daten und mit den transformierten Daten aus (a) und (b) (Befehl `regress`). Abhängige Variable ist die Hausarbeit (`hw`), unabhängige Variablen sind:

- Monatsverdienst
- HH-Einkommen
- Alter
- Dummy verheiratet

Beschränken Sie den Datensatz jeweils auf die Personen, die mindestens 2 gültige Werte für alle Variablen haben.

Wie unterscheiden sich die Ergebnisse?

(Hinweis: Beim Demeaning können Sie so vorgehen: Berechnen Sie zuerst den Mittelwert einer Variablen `mvarname` und speichern ihn in einer neuen Variablen (Befehl `egen`). Dann erstellen Sie eine neue Variable `dmvarname` für die transformierten Werte. Verwenden Sie eine Schleife, um diese Operation für alle Variablen durchzuführen.)

4. Schicken Sie Ihr Do-file an [gerrit.bauer@lmu.de](mailto:gerrit.bauer@lmu.de). Geben Sie im Betreff der email das Kürzel „MVS017IhrNachname“ an.

## Übung Multivariate Analyseverfahren – SoSe 2017

Dr. Gerrit Bauer

### Übungsblatt 4 – Abgabe bis 28.06.2017

Wir arbeiten mit dem Paneldatensatz `housework.dta`.

1. Erstellen Sie ausgehend von Variable `marstat` zwei neue Variablen:
  - Variable `marry` ist eine Dummy-Variable, die gleich 1 ist, wenn eine Person derzeit verheiratet ist, und sonst 0.
  - Variable `ever_marry` ist eine Dummy-Variable, die gleich 1 ist, falls eine Person zu irgendeinem Zeitpunkt verheiratet ist, und sonst 0.
2. Beschränken Sie den Datensatz auf Personen mit mindestens 2 gültigen Werten auf den Variablen `hw labgro marstat hhinc age` (s. Do-file `05_Übungsblatt3.do`). Verwenden Sie `xt` Befehle, um folgende Fragen zu beantworten:
  - Wie viele Personen und Personenjahre beobachten wir? Wie viele Personenjahre beobachten wir im Mittel für die Personen?
  - Wie viele Personen heiraten zum ersten Mal, während wir sie im SOEP beobachten?
  - Wie viele Personen werden jemals in einer Ehe beobachtet?
3. Erzeugen Sie eine Variable `yrsmarried`. Die Variable zählt für jede Person die Jahre, die vergangen sind, seit die Person zum ersten Mal in einer Ehe beobachtet wird (`marry=1`). Überlegen Sie sich, für welche Personen/Personenjahre der Wert für `yrsmarried` der Ehedauer entspricht.
4. Schicken Sie Ihr Do-file an [gerrit.bauer@lmu.de](mailto:gerrit.bauer@lmu.de). Geben Sie im Betreff der email das Kürzel „MVSO17IhrNachname“ an.

## Übung Multivariate Analyseverfahren – SoSe 2017

Dr. Gerrit Bauer

### Übungsblatt 5 – Abgabe bis 05.07.2016

Wir konstruieren eine Stichprobe aus dem SOEP, um die Hausarbeit von Frauen (ihren Zeitaufwand) im Lebensverlauf zu analysieren. Gehen Sie aus von dem Paneldatensatz `housework.dta`.

1. Schränken Sie die Stichprobe ein. Erzeugen Sie eine Variable `sample`, die zunächst gleich 1 ist. Schließen Sie dann sukzessive aus:
  - Männer (Setzen Sie Variable `sample` auf den Wert 0)
  - Personenjahre mit fehlendem Wert auf Variable `hw` (`sample=-1`)
  - Personenjahre mit fehlendem Wert auf mindestens einer der unabhängigen Variablen `marstat kids14 hhinc age lfs2`. (`sample=-2`)
  - Personen mit weniger als 2 Personenjahren (`sample=-3`)

Schauen Sie sich die Verteilung von `sample` an und interpretieren Sie sie.

(Hinweis: „Sukzessive“ heißt, dass Sie das Ausschlusskriterium jeweils auf die verbleibende Stichprobe aus dem vorherigen Schritt anwenden.)

2. Stellen Sie die Stichprobenziehung in einer Word-Tabelle dar (oder anderes Textverarbeitungsprogramm). Verwenden Sie dazu die Befehle `estpost` und `estout` (Sie müssen dafür ein Ado-Paket von Ben Jann installieren, Sie finden es mit `net search estout`).
3. (\*) Erstellen Sie eine Variable `sample2`, die zunächst gleich 1 ist. Wenden Sie die Ausschlusskriterien aus Aufgabe 1. nur auf der Personenebene an. Schließen Sie sukzessive aus:
  - Männer (`sample2=0`)
  - Personen, die für `hw` immer fehlende Werte haben (`sample2=-1`)
  - Personen, die immer fehlende Werte für mind. eine der unabhängigen Variablen `marstat kids14 hhinc age lfs2` haben. (`sample=-2`)
  - Personen mit weniger als 2 Personenjahren (`sample=-3`)

Stellen Sie die Stichprobenziehung auf Personenebene in einer Word-Tabelle dar.

4. Schicken Sie Ihr Do-file und die Tabelle aus Aufgabe 2. an [gerrit.bauer@lmu.de](mailto:gerrit.bauer@lmu.de). Geben Sie im Betreff der email das Kürzel „MVSO17IhrNachname“ an.

(\*) Zusatzaufgabe (nicht verpflichtend)

## Übung Multivariate Analyseverfahren – SoSe 2017

Dr. Gerrit Bauer

### Übungsblatt 6 – Abgabe bis 12.07.2017

Wir untersuchen den Effekt der Heirat auf die Hausarbeit von Frauen (ihren Zeitaufwand). Gehen Sie aus von dem Paneldatensatz `housework.dta`.

1. Schränken Sie die Stichprobe ein. Schließen Sie aus:
  - Männer
  - Personenjahre mit fehlendem Wert auf Variable `hw`
  - Personenjahre mit fehlendem Wert auf mindestens einer der unabhängigen Variablen `marstat kids14 hhinc age lfs2`
  - Personen, die nie im Zustand „ledig“ beobachtet werden
  - Personenjahre nach der 1. Ehe (d.h. ab dem Zeitpunkt, ab dem eine Person zum ersten Mal getrennt/geschieden oder verwitwet ist)
  - Personen mit weniger als 2 Personenjahren

(Hinweis: Gehen Sie aus von Do-File `07_Übungsblatt5`.)

2. Schätzen Sie den Effekt der Heirat auf die Zeit für Hausarbeit mit folgenden statistischen Modellen.
  - Between-Effects
  - Pooled OLS
  - Random-Effects
  - Fixed-Effects

Kontrollieren Sie in den Modellen für die Zahl der Kinder im Haushalt (`kids14`, 3 Kategorien (Ref. kein Kind): 1 Kind, 2 Kinder, 3 oder mehr Kinder) und das Haushaltseinkommen (`hhinc`) und das Erhebungsjahr (`year`, Dummies für einzelne Jahre, Referenz 1992).

3. Vergleichen Sie die Ergebnisse aus Aufgabe 2. Für welches Modell entscheiden Sie sich?
4. (\*) Schätzen Sie den Effekt der Heirat mit einer *Dummy Impact Function* (D.h.: Sie spezifizieren einen zeitveränderlichen Effekt der Ehedauer. Zur Konstruktion der Ehedauer, s. Do-File `06_Übungsblatt4`). Wie verändert sich der Effekt der Heirat im Verlauf der Ehe?
5. Schicken Sie Ihr Do-file an [gerrit.bauer@lmu.de](mailto:gerrit.bauer@lmu.de). Geben Sie im Betreff der email das Kürzel „MVSO17IhrNachname“ an.

(\*) Zusatzaufgabe (nicht verpflichtend)



## Übung Multivariate Analyseverfahren – SoSe 2017

Dr. Gerrit Bauer

### Übungsblatt 7 – Abgabe bis 26.07.2017

Wir schätzen Wachstumskurven-Modelle für die Hausarbeit von Frauen (ihren Zeitaufwand). Gehen Sie aus von dem Paneldatensatz `housework.dta`.

1. Schränken Sie die Stichprobe ein. Schließen Sie aus:
  - Männer
  - Personenjahre mit fehlendem Wert auf Variable `hw`
  - Personenjahre mit fehlendem Wert auf mindestens einer der unabhängigen Variablen `marstat kids14 hhinc age lfs2`
  - Personen mit weniger als 2 Personenjahren

(Hinweis: Gehen Sie aus von Do-File `07_Übungsblatt5.`)

2. Schätzen Sie den Alterseffekt mit Fixed-Effects (Wachstumskurven-) Modellen. Kontrollieren Sie Periodeneffekte. Spezifizieren Sie den Alterseffekt
  - parametrisch (linearer, quadratischer, kubischer Term)
  - mit Dummy-Variablen

Interpretieren Sie die Ergebnisse.

3. Wie erklären Sie sich den Altersverlauf aus Aufgabe 2.? Schätzen Sie erweiterte Modelle, um diese Frage zu beantworten. Berücksichtigen Sie Familienstand, Zahl der Kinder im Haushalt, Erwerbsstatus und Haushaltseinkommen als zusätzliche unabhängige Variablen.
4. (\*) Schätzen Sie den Alterseffekt nun für Ost- und Westdeutschland (Variable `east`). Schränken Sie dazu die Stichprobe weiter ein auf Personen, die nur in Ost- bzw. nur in Westdeutschland leben (`east` wird dadurch zeitkonstant). Nehmen Sie an, dass Periodeneffekte in Ost und West gleich sind (ebenso die Effekte der weiteren unabhängigen Variablen). Prüfen Sie die These, dass der Aufwand für Hausarbeit westdeutscher Frauen im Alter von 26-45 Jahren (verglichen mit der Altersgruppe 16-25 Jahre) stärker steigt als dies für ostdeutsche Frauen der Fall ist. Prüfen Sie, ob sich dieser Unterschied damit erklären lässt, dass sich die Familiensituation und das Erwerbsverhalten in Westdeutschland in dieser Phase des Lebensverlaufs stärker ändern.
5. Schicken Sie Ihr Do-file an [gerrit.bauer@lmu.de](mailto:gerrit.bauer@lmu.de). Geben Sie im Betreff der email das Kürzel „MVSO17IhrNachname“ an.

(\*) Zusatzaufgabe (nicht verpflichtend)

## Übung Multivariate Analyseverfahren – SoSe 2017

Dr. Gerrit Bauer

### Übungsblatt 8 – Bearbeitung in der letzten Sitzung

Wir schätzen Wachstumskurven-Modelle für die Hausarbeit von Frauen (ihren Zeitaufwand). Gehen Sie aus von dem Paneldatensatz `housework.dta` und Do-file `09_Übungsblatt7.do`. Beschränken Sie die Stichprobe auf Frauen mit mindestens 2 Personenjahren mit gültigen Werten auf den Variablen `hw` `marstat` `kids14` `hhinc` `age` `lfs2`.

1. Schätzen Sie den Alterseffekt mit Fixed-Effects (Wachstumskurven-) Modellen. Kontrollieren Sie Periodeneffekte. Spezifizieren Sie den Alterseffekt
  - parametrisch (linearer, quadratischer, kubischer Term)
  - mit Dummy-Variablen

Interpretieren Sie die Ergebnisse. Erstellen Sie dazu anhand der Regressionsergebnisse für jedes der beiden Modelle einen Profile Plot der vorhergesagten Werte in Abhängigkeit des Alters. Verwenden Sie `margins` und `marginsplot`. (Speichern Sie die beiden Grafiken zuerst mit `graph save`; verwenden Sie `graph combine`, um beide Grafiken dann in einer Abbildung darzustellen.)

2. Schätzen Sie den Alterseffekt nun für Ost- und Westdeutschland (Variable `east`). Schränken Sie dazu die Stichprobe weiter ein auf Personen, die nur in Ost- bzw. nur in Westdeutschland leben (`east` wird dadurch zeitkonstant). Nehmen Sie an, dass Periodeneffekte in Ost und West gleich sind (ebenso die Effekte der weiteren unabhängigen Variablen).

Untersuchen Sie den Unterschied in der Zeitverwendung für Hausarbeit zwischen ost- und westdeutschen Frauen im Lebensverlauf. Stellen Sie die Ergebnisse grafisch mit einem Profile Plot und mit einem Conditional Effect Plot dar. Verwenden Sie `graph save` und `graph combine`, um die beiden Grafiken in einer Abbildung darzustellen.

(Hinweis: Überlegen Sie sich zuerst, welches Modell (FE oder RE) geeignet ist, und wie Sie mit dem APC-Problem umgehen.)